DOCUMENT RESUME

ED 171 799 TM 009 506

AUTHOR Hecht, James T.
TITLE The Development and Structure of Professional Examinations Planned for National Use.
PUB DATE Apr 79
NOTE 27p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Certification; *Equivalency Tests; National Programs; Personnel Evaluation; Professional Occupations; *Professional Personnel; Scoring; *Standards; *Test Construction; Testing

ABSTRACT

The process typically employed by testing services in developing professional tests for national use is described: (1) determination of professional standards; (2) development of test specifications; (3) test construction; (4) test registration and administration; and (5) scoring, analysis, and reporting. To determine professional standards, input should be obtained from members of the profession; adequate minority group opinions solicited; standards refined; and discussions leading to consensus encouraged. Considerations in developing test specifications include the purpose of the examination; the nature of the test results, number of test items and time needed to administer the test; type of test items; the definitions of content and skill domains; test item characteristics, such as format, difficulty, and discrimination; test reliability; and test validity--including face validity, content validity, and criterion related validity. Steps in developing test specifications include preparing a preliminary test outline; initiating a test specification workshop; and following up the workshop with clarification and refinement of specifications. Test construction involves test item development; field testing; test assembly; and printing. Test registration and administration involves registering examinees; establishing test centers; and administering the test. The final step involves scoring, analyses, and reporting of results. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED171799

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper Presented to a Meeting of the
National Council on Measurement in Education
April 1979

The Development and Structure of Professional Examinations Planned for National Use

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James T. Hecht

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

By:

James T. Hecht, Ph.D., M.P.H.
Director of Professional Assessment Programs
Research and Development Division
American College Testing Program
Iowa City, Iowa

TM009 506

ABSTRACT

The purpose of this paper is to describe the process that is typically employed by testing services to develop professional examinations planned for national use. Since this paper is part of a symposium that includes other discussions on various aspects of examinations used for licensure and/or certification, an all-inclusive review was not prepared. Further, as there are many acceptable routes to examination development, this paper does not contain the perfect, ideal or most complete approach.

The major steps in the development and implementation of professional examinations planned for national use are reviewed and are as follows:

1. Determination of professional standards
2. Development of test specifications
3. Test construction
4. Test registration and administration
5. Scoring, analysis, and reporting

INTRODUCTION

Consistent with the structure of the overall symposium topic of "Perspectives of Licensure Examinations: National vs. Local" and within the parameters of the paper topic of "The Development and Structure of Professional Examinations Planned for National Use," the key procedures for preparing and implementing national professional examinations will be presented. However, since an all inclusive discussion of professional examinations cannot adequately be treated within the confines of this short paper, the scope of the paper's topic will be limited to the application of test development and psychometric procedures utilized in typical professional assessment programs. In addition, three general assumptions will guide the tone of the paper: (1) while commonalities exist across most national professional examinations, no two professions and/or occupations require precisely the same degree or level of sophistication of psychometric input; (2) the quantities of human resources available within a profession will influence the capability of that profession to apply psychometric theory; (3) there are alternative practical psychometric approaches for accommodating unique professional assessment requirements. Thus, this paper does not purport to offer the perfect approach for developing and implementing national professional assessment examinations.

Professional associations and/or governmental commissions are usually charged with the responsibility for certification and/or licensure programs. The staff configurations and other resources of such groups will influence the magnitude of the potential interactions that could occur between a given group and a testing service. Although some non-testing service organizations employ persons holding measurement credentials and command natural resources that allow the accomplishment of accepted psychometric activities, most such organizations do not have immediate in-house access to measurement expertise.

In fact, a professional organization or governmental commission that can successfully operate a national professional assessment program is, indeed, exceptional. For most of these organizations the working relationships that exist with testing services constitute essential components of their program's credibility.

The preparation of a professional examination program for use on a national basis is a complex task, requiring the coordination and management of many unique and independent resources. Professionals holding doctoral degrees in the fields of educational measurement and statistics usually constitute the basic staff resource in such activities, and are generally responsible for conceptualizing both the structure of the measuring instrument and the theoretical premises upon which the examination results are based. Computer resources that represent the most current technology and the most advanced computer thinking are needed to handle the processing of measurement data. Printed materials must be produced through precision quality control and under strictly secure conditions. And, above all, objectivity and credibility must be maintained. It is the collection and articulation of such capabilities that allow professional testing services to utilize valuable resources effectively.

The development and implementation of professional examinations planned for national use typically requires the following major steps:

(1) Determination of professional standards
(2) Development of test specifications
(3) Test construction
(4) Test registration and administration
(5) Scoring, analysis, and reporting

Although all five of the above referenced steps will be addressed, steps 1 and 2 will receive the most attention. This paper will conclude with a summary of the implementation steps, as applied to national examinations that contain local components.

DETERMINATION OF PROFESSIONAL STANDARDS

While ongoing certification and/or licensure programs have usually already experienced job analyses and/or program objectives surveys, or established standards of practice, for most new programs the determination of professional standards constitutes the initial step in program development.

To illustrate this step, this section describes a set of typical plans for defining standards of competence for a newly developed field. An overview of a plan for defining standards of competence is presented first, followed by a review of the implementation steps.

Overview of a Plan for Defining Professional Standards

In formulating procedures for defining professional standards, it is important to include provisions for obtaining:

(1) substantive input from the associations of the profession and their members;

(2) adequate representation of minority group members;

(3) open-ended constituency input early on in the consensus gathering process that can bring increasing specificity to the standards with each new step in the process; and

(4) professional association input that leads to interactive discussions on varying viewpoints and eventually to consensus.

Early in the process, the testing service reviews and elaborates on tasks (1 - 4 above) to be completed. In general, however, such a plan would call for a series of regional meetings involving selected members of the constituency. Specifically, a number of meetings comprised of small groups of members should be held in different regions of the country for the purpose of obtaining preliminary statements of standards of practice.

The chairpersons of the regional meetings should plan to meet as a single group with the testing service staff members prior to the regional meetings. The testing service assumes primary responsibility for planning the key aspects of this preliminary meeting with the chairpersons. The purpose of this meeting is to plot a strategy and structure for the regional meetings. An important outcome should be the preparation of materials for advance mailing to potential participants of the regional meetings. Those materials typically include information about the purpose of the regional meetings, the agenda, and work assignments which should be carried out prior to the regional meetings.

Subsequent to the meetings, each chairperson assumes responsibility for preparing a position paper which synthesizes the views of participants at the meeting. These position papers are forwarded to the testing service where papers from all of the regional meetings are synthesized, areas of agreement outlined, and issues needing resolution identified. Once this task is completed, copies of the resulting materials are mailed to an Advisory Panel (additional content experts in the profession), specified officers of the professional association, and the chairpersons of each regional meeting. Collectively, these individuals are assembled as a National Task Force for the purpose of completing a single Task Force report which includes recommendations for establishing professional standards. These recommendations and other issues addressed at the regional meetings and the National Task Force Meeting are then incorporated into a final report.

Specific Steps in Specifying Standards

While the preceding discussion provided a capsule overview of the steps involved in defining standards of competence for professional practice, this part provides a specific breakdown on each of the major steps in that process.

It is worth noting at this point that some aspects of the testing service plan include arbitrary decisions (e.g., the number of regional meetings, the makeup of the National Task Force), which could easily be modified to reflect input from the professional association.

1. Plan and implement regional meetings
   a. Define geographical region boundaries
   b. Select regional meeting representation characteristics
   c. Select dates and sites
   d. Orient regional meeting chairpersons
   e. Hold regional meetings and produce position papers

2. Plan and implement National Task Force Meetings
   a. Synthesize regional position papers
   b. Prepare draft of standards

After receiving and incorporating suggestions from members of the National Task Force into the statements of standards, the testing service incorporates them into its final report. At this point a set of professional standards exists.

DEVELOPMENT OF TEST SPECIFICATIONS

For a number of professional associations and/or licensure commissions, development of test specifications is an entry point into the examination preparation cycle. These groups have already developed standards through task analysis, statutes preparation and/or job analyses. Other groups have arrived at these standards through the accomplishment of steps described in the previous section on standards. In either case, some form of standards must be available before the test specifications step can be conducted.

This section describes the procedures that are recommended for determining the scope and content of an examination for a given professional group. Collectively, the attributes of the examination are referred to as specifications or blueprints, and they are crucial to the construction of a valid and reliable testing program.

> The term "specifications" refers to the blueprints for the examination, that is, the guidelines which the test developer uses to build the examination. Just as a blueprint is utilized in building a physical structure, likewise a test blueprint is used to construct a test. Developing a test involves a number of sequential steps, each incrementing on the results of previous steps and each, in turn, serving as the foundation for later decisions and actions. Developing a test blueprint is a first step and is crucial to the preparation and development of test items. Precise blueprints are conducive both to the avoidance of pitfalls in the test development process and to the assurance of valid instrumentation.

The discussion of test specifications in this section is divided into two parts: the first deals with general and technical considerations in the design of an examination; and the second describes the actual steps that the testing service and the professional association execute in the process of preparing an operational set of test specifications.

General and Technical Considerations

Included under the rubric of general and technical considerations are the following topics: (1) the statement of the purpose(s) of the proposed examination, (2) the nature of the examination results or scores, (3) the examination length in both time and items, (4) the type of items to be included in the examination, (5) a definition of the content domain, (6) a definition of the skills domain, (7) the characteristics of the items, (8) the test reliability, and (9) the test validity.

Purpose(s) of the Examination

An important consideration in the development of any examination is the use(s) to be made of the examination. For example, if an examination is to be used to select an individual who has more skills than other individuals, a norm-referenced test (one permitting comparisons across individuals) may be called for. On the other hand, if an examination is to be used primarily to provide individuals with diagnostic information about their areas of strengths and weaknesses, a criterion-referenced test may be needed. In the latter case, it is usually unimportant how one individual performed in relation to others. In either case, or in the case of some combination

of the two, the resultant examination must be fair to all individuals. That is, all qualified individuals should be able to pass the examination.

Examination Results

The results of the examination should indicate that a successful candidate has met or exceeds the minimum criteria established by the professional association for recognition as a professional. This would indicate to the public that the individual holding such a credential has obtained a level of competence worthy of inspiring public trust.

Examination Length

The number of items included in the test must be sufficient to adequately sample the specified content and skills domain. Typically, the number of items in a test is also influenced by item types to be used and expert judgments about the amount of time the average candidate should take in responding to a question.

The time allotted to an examination should be adequate to reliably sample appropriate candidate behaviors. It should be reasonable without inducing great fatigue. The examination length in terms of time and number of items is dependent upon the collection of content and skill levels required for successful completion. The more complex the content and skills, the longer the time needed for each item and the fewer the number of items that can be administered within a fixed time period.

Type of Examination Items

In general, there are two major types of examination items--the essay type and the objective type--both having their advantages and disadvantages. In a lengthy examination with a large number of candidates participating, it would be difficult if not impossible, to cope with essay type responses. Good essay type items are difficult to construct and are much more difficult

to score. The objective type examination, consisting of multiple-choice, true-false, or matching items, is preferred for large numbers of candidates and is satisfactory for measuring important knowledge and skills. The objective type examination has a further advantage in that it can be easily machine scored.

A preliminary review of the content and skills domains pertinent to the work of the professional would probably suggest that the examination consist of multiple-choice items which could be used effectively to assess those domains. The examination could contain some charts and graphs along with paragraphs on specific topics for testing comprehension. While candidates might be required to work through problems or fill out forms, actual scoring would be via responses to multiple-choice questions that relate to the problems.

Content Domain

Defining the scope and content of the examination is the most critical and detailed process preceding item developemnt. (Preliminary work of this nature would already have been accomplished through the standards development procedures reviewed in the previous section.) The important steps to be covered would include (1) checking the significant topics against the content outline and (2) conducting analyses of the weights assigned to the topics with respect to appropriateness to a relevant criterion.

Skill Domain

A second and equally important process in developing the examination blueprint is to define the level of skill to be measured. Stated another way, the question to be addressed is, what should the candidates be able to do with content? The level of skills can be viewed as falling on a continuum

from simple memory to complex problem solving as exemplified by the six categories listed below:

1.	Recall	The building blocks for all later knowledge; essentially a memory measurement, i.e., what has been seen or heard. Define, list, and name.
2.	Comprehension	Translate and interpret transfer of meaning from one mode of presentation to another, e.g., graphs, pictures, tables. Restate, recognize, express.
3.	Application	Apply knowledge to a problem situation, i.e., candidate must determine the necessary operations and perform them. Formulas--compute and answer or interpret a problem and the application of principles to find a solution.
4.	Analysis	A process used in problem solving. Compare and contrast a situation to derive a conclusion or generalization; use of two or more variables to determine their respective impacts on a third variable.
5.	Synthesis	Solving a problem with more than one potential answer; construct, compose, propose, or create a plan.
6.	Evaluation	Developing one's own standards or using someone else's to evaluate procedures, hypotheses, or conclusions; judge, rate, or appraise some object with respect to the standards.

It is not necessary for a given examination to include all of the skill levels across the content areas. However, that which is determined to be relevant and appropriate to the objectives of the examination will influence the particular skill levels sampled by the test. After determining which skills are important across the content domain a weight is assigned to each of the skills to decide the proportion of the test that will be focused on each skill. The combination of content and skill domains results in a test specifications matrix or blueprint. An example of a test blueprint is provided in Figure 1.

Note: The six categories above were adapted from Bloom, Taxonomy of Educational Objectives: Handbook I. The Cognitive Domain, New York: David McKay, 1956.

Figure 1

Test Blueprint "Shell" for a Professional Examination

Content	Skills						Weights
	Recall	Comprehension	Application	Analysis	Synthesis	Evaluation	Row Totals
I. Category							(XX)
A.							XX
B. (Subcategories)							XX
C.							XX
D.							XX
II. Category							(XX)
A.							XX
B. (Subcategories)							XX
C.							XX
D.							XX
III. Category							(XX)
A.							XX
B. (Subcategories)							XX
C.							XX
D.							XX
Column Totals	XX	XX	XX	XX	XX	XX	

Item Characteristics

There are three important aspects of a test item that should be determined in the test specifications phase of the test development process: item format; item difficulty; and level of item discrimination. A description of each of these item characteristics follows.

Item Format

Considerations regarding the format of the items result from decisions made regarding the best method to present the test items to the candidate (written tests vs. performance tests), time factors (group vs. individual administration), and cost factors. Usually the costs of individual performance testing are higher than the costs of group and written tests.

Item Difficulty

The term "difficulty" as used in test development is somewhat of a misnomer in that the difficulty index refers to the proportion of candidates correctly answering a particular test item. Thus a high difficulty index reflects an item which proved relatively easy for the group of individuals whose performance is being analyzed.

When performance on all items of the test is analyzed, one can derive a mean item difficulty for that test, this value being equal to the arithmetic average of the proportions of candidates responding correctly to the various items within the test. To determine a desirable value of this average for the examination, it is necessary to consider the purposes that the examination is intended to serve.

Level of Item Discrimination

The item discrimination index represents the extent to which the various test items successfully differentiate between candidates of high ability and those of low ability (ability being used synonymously with either high or low performance on the test as a whole). Thus, the items on one section of the test would be said to discriminate poorly if, on the average, they were answered correctly by candidates doing poorly on the test as a whole, and incorrectly by candidates doing very well on the test as a whole.

Reliability

Reliability indicates how much confidence can be placed in the test results. In the case of the professional examination, if the professional association chooses a norm-referenced approach, two computation procedures could be used to verify test reliability. First, the correlation between

the odd-numbered and even-numbered items on the test (corrected by the Spearman-Brown Prophecy Formula) would be used as one indication of the precision of measurement. This technique is best known as odd-even or split-half reliability. Second, the Kuder-Richardson Formula 20 (KR_{20}) indices should serve to indicate the degree of reliability of the examination.

Should the professional association choose a criterion-referenced approach to testing, the major reliability concern is that repeated administrations of the test to the same candidates result in identical pass/fail decisions. Thus, in this case, a test-retest reliability could be computed.

Validity

Validity refers to the extent to which the results of an instrument serve the particular uses for which it is intended. In other words, does the test measure what it is supposed to measure? The professional examination should measure minimum competencies required on the job. There are three types of validity that are important to the professional examination: face validity, content validity, and criterion-related or predictive validity. A brief description of each of these types of validity follows.

Face Validity

The professional examination has face validity to the degree that it has credibility with the professionals themselves and the public they serve. The tests should not only measure important skills, but have the appearance of doing so. Face validity could be achieved by thorough content coverage and a display of sufficient numbers of items to assess that content.

Content Validity

A test can be said to have content validity to the degree that it measures the competencies it is designed to measure. To assure that the test has adequate content validity, the testing service attempts to make certain that every item provides a measure of the types of competencies deemed important for the profession. Moreover, the testing service takes care that a proper balance of items representative of each of the required competencies is included in the test.

Content validity can, in part, be accomplished by carefully defining the context in which the competencies are to be measured, the various content areas to be tested, and the proportion of items required for each combination of skill and content area. Precision of specifications to item writers is regarded by the testing service as crucial to the construction of tests which do, in fact, measure what they are intended to measure.

Criterion-related Validity

This is the degree to which test performance is related to some other valued measure of performance. The professional examination has criterion-related validity if it adequately identifies successful and unsuccessful candidates and if it adequately estimates their competencies. Because the results of the professional examination could have far-reaching consequences, it is recommended that a selected representative sample of practicing professionals be used as a pilot group to ascertain the validity of the instrument as a predictor of job performance.

Procedures for Developing Test Specifications

The practical steps involved in producing quality test specifications are as follows: preparation of a preliminary test blueprint based on the standards of practice; execution of a test specifications workshop; and completion of test specifications workshop follow-up activities. The following discussions address the topics cited above and describe in practical terms the procedures followed to implement the work steps.

Preliminary Preparation of a Test Blueprint

The following steps are recommended in developing a preliminary outline for the professional examination. First, the professional association should identify a task force of content experts to guide and evaluate the development of the examination. Second, the testing service

should prepare a preliminary outline based upon the professional standards. Third, the task force and the testing service should jointly define the major categories of content to be evaluated. Fourth, the subcategories of content for each of the major categories should be determined and specified.

Test Specifications Workshop

Members of the testing service typically work cooperatively with the task force to identify the specific subcategory content. Next, the assignment of relative weights or proportions to each of the major categories and subcategories of the content based on some external criterion is determined by members of the task force and the testing service. An example of a criterion is the level of importance of the category to competent and professional practice. This is determined by reviewing the professional standards.

Test Specifications Workshop Follow-up

The testing service staff continues to process materials developed at the workshop after the task force has returned home. Clarification and refinement of the test specifications is accomplished through follow-up telephone conversations and correspondence. Eventually a workable test specifications matrix results.

TEST CONSTRUCTION

Once the table of specifications for the examination has been developed and approved by appropriate professional association representatives, the translation of these blueprints into the desired examination begins.

This process involves a number of steps, each designed to ensure the time and cost-effective development of test materials while at the same time guaranteeing that the developed materials are as valid as possible for the stated purpose of the examination.

Item Development

Examination construction typically follows a traditional pattern of combining content expertise (professional association) with measurement expertise (testing service). Item writer stimulus materials, which describe a variety of item formats and provide general information and advice related to the construction of technically sound items are usually prepared by the testing service staff. The manner in which item writers are trained and items are written, depends upon the professional association's degree of involvement in the item writing process. The discussion that follows assumes a moderate degree of participation by the professional association in the item development process.

The professional assocations identifies a number of persons from within the professional field who are willing to develop small units of test questions. The next step consists of conducting an item writer's workshop in which content experts are trained to be item writers. A portion of the actual items are written at the item writer's workshop, but the majority of the item development is carried out after the workshop with follow-up editing and review by mail.

The typical test development process stresses the use of many content experts to write examination questions. This, however, necessitates that special security procedures be enforced. For example, all

communication between the testing service and the content experts/item writers is conducted either by telephone or by first-class mail. In these communications, item writers are given both directions regarding sound item writer practices and directions on general security concerns. As a major component of this latter set of directions, they are instructed to keep all items under development in a locked file or cabinet to prevent unauthorized access to these items.

After the items have been constructed, the items are thoroughly reviewed by a group designated by the professional association to ensure that each item is: (1) content accurate; (2) relevant to the table of specifications; and (3) testing a significant concept and skill. Items failing to meet these and other criteria are eliminated from the item pool or revised.

In addition to this review, the testing service editorial staff reviews the items to ensure that each item is (1) clearly stated; (2) grammatically sound; (3) free of clues to other items; and (4) unbiased in its language and meaning. Once again, any item failing to meet any of these criteria is either revised or eliminated from the item pool.

Field Testing

Items developed by even the most experienced and skilled item writers are not always adequate for use in a professional examination. Even after close scrutiny by psychometric and content specialists, certain items that appear technically sound may fail to perform as expected when administered. For example, items may prove to be unclear or confounding to a large number of candidates, causing them to respond incorrectly,

though they are highly knowledgeable in a specific content area. Therefore, the construction of the best possible version of a test demands that all items be field tested and those items that fail to function effectively be identified and either revised or eliminated from the item pool before the final versions of the examination are assembled.

Test Assembly

Following field testing, the forms of the examination to be administered are assembled according to the table of specifications approved by the professional association. To ensure score comparability from test date to test date a certain portion (usually one-fourth to one-third) of the total number of items in a given examination are selected for reuse in future examinations. These common items are used to equate the scores from one examination administration to the next. Once the tests are assembled, the testing service editorial staff reviews the materials for clarity and fairness with consideration for sex and race. At that time, the tests are submitted for final approval by the professional association. This review serves as the final check on content accuracy and item keying prior to the administration of the tests.

Test Booklet Printing

Depending upon the quantity and complexity of the printing demands associated with tests and support materials for the professional assessment program, the testing service either uses its own in-house printing facilities or else contracts with outside commercial printing concerns. In the case of in-house printing, security is relative easy to ensure, since buildings are guarded during nonbusiness hours and all staff having

access to materials during each stage of production have signed pledges to preserve the confidentiality of all materials with which they come in contact. This understanding on the part of all staff, coupled with standard building security practices during nonbusiness hours works exceedingly well. When materials must be sent to commercial printers, strict rules govern the procedures to be followed.

TEST REGISTRATION AND ADMINISTRATION

For reasons of logistics, economies of scale, and experience, the testing service is in an excellent position to provide efficient and effective registration and administration services. The national network of test centers and personnel that are an integral part of the testing service structure can be extremely difficult for a professional association to duplicate. To illustrate this perspective, this section will address the key operational aspects of a typical professional examination planned for national use. The three critical phases of this activity are as follows: registration, test center establishment, and test administration.

Registration

Registration constitutes the staging of the testing process. The testing service can perform all or part of the activities related to the registration of applicants for the testing program. The activities generally involve preparing, distributing, and processing registration materials. Typical steps in the process consist of the following: (1) designing and preparing printed registration materials; (2) developing and maintaining the mailing list for registration materials distribution;

(3) distributing and retrieving registration materials; and (4) issuing admission tickets. The extensive data collection and information reporting experience of a testing service can be brought to bare on the national testing program to develop and implement the most efficient registration system possible.

Test Center Establishment

Although establishment of test centers is a complex task, testing services have developed efficient systems for handling the process. Because of the volume of activities conducted at the test center level, it is common for the testing service to maintain hundreds of testing centers nation-wide.

The testing service can provide test centers in each state or geographic region where the number of candidates warrants. It is desirable, however,that at least 20-25 candidates be assigned to each center. It is customary to insure that the majority of candiates do not have to travel long distances to reach a given test center. Special testing arrangements can also be made for physically handicapped persons.

The testing service selects the test personnel from test supervisors who administer other tests for the testing service. These persons are highly experienced in the art of proctoring and understand the need for maintaining standardized testing conditions. For new persons who enter this field, the testing service provides an organized series of training sessions. The bottom line is that the professional association can expect the testing service to assign well-qualified proctors to adminster the test.

Test Administration

While there are many services performed under the category of Test Administration, a number of customary activities are as follows:

(1) prepare and print test administration documents; (2) develop procedures for administering the test and implementing test security measures; (3) distribute test materials; (4) administer the test; (5) receive and inventory returned test materials; (6) follow up on all test security matters; and (7) store test materials.

SCORING, ANALYSES, AND REPORTING

The testing service employs psychometric experts that are proficient in balancing the interface between the computer world and the understandable interpretations of data that most professional associations can utilize. It is highly recommended that the testing service handle this phase of the project's activities.

Examination Scoring

After each test date, candidates' responses to the test items are recorded on computer tape using a high speed optical scanner. This recording of candidates' responses has become the standard in commercial test scoring. This rapid scanning process allows the scoring of results in a most expeditious manner.

Examination Analyses

Two types of statistical analyses are customarily needed following each examination administration: (1) analyses reflecting summary statistics for the examination and (2) analyses for each item.

In general, the principal purposes of the analyses are to provide descriptive information about the examinations and to provide evidence to support the reliability and validity of the examinations. This is not to say that these analyses are the sole basis for judging examination reliability

or validity. For example, for the national professional examinations, the most important type of validity is content validity. For the most part, content validity will be established not through statistical procedures but through careful procedures for test specifications, item construction, and examination review by subject matter specialists.

Reporting of Results

The professional association needs to determine the type of information to be sent to successful and unsuccessful candidates. The testing service is in an excellent position to assist in this matter as it customarily prepares numerous types of reports such as: candidate score reports, analyses of performance trends, item analyses, and other statistical information.

SUMMARY

The discussions contained in the previous sections offer an approach for developing and implementing a professional assessment examination program planned for national use. In practical terms, however, many of the steps and activities could be altered and/or expanded to meet the needs of the particular discipline and profession under consideration. To focus this paper more specifically on the real estate profession, cited by some of the other symposium presenters, the aspects of preparing a national program containing local components are addressed.

National Standards

The standards step is usually at least partially completed, by virtue of the statutes that are in force at the local (state) level in an occupation such as real estate. While refinements to the process can be achieved through constituent surveys and job analysis techniques, the statutes provide the

basis of the examination planning activities. All of the state statutes represent standards. The theoretical national standards emerge as the commonalities between the various state statutes.

Test Specifications

The concept of a national test specifications matrix may be difficult to fully implement as a practical step. It is possible to prepare a consensus document, formulated through input from real estate educators and commission representation via meetings and draft refinements. However, to actually apply the theoretical national test specification matrix as is may not be possible. Each state should have input into the matrix weights prior to implementation and this process may produce a minute difference between the state common matrix and the theoretical national matrix. In theory, however, the common matrix appears to be a plausible approach.

Construction of Items

To develop the required bank of items, where possible, the testing service can employ two procedures. Testing service staff members who are expert in item construction techniques can prepare test items based on the common statutes. These items need several content review steps. A second procedure which the testing service could employ in developing test items involves contracting with prominent educators to construct items. Under that type of arrangement, qualified persons knowledgeable in the content and skills crucial to the roles of practitioners play a role in the development process--thus introducing an additional element of content validity to the examinations.

Integrating the National Examination with Local Supplements

The integration process varies according to the needs of individual states. Since local supplements are designed to be state-specific, procedures for merging state items with the national examinations could vary. The final products, however, could very well be alike from a structural perspective (e.g., 100 national items, 25 local items).

Conclusion

The notion of national examinations (or components, thereof) can more realistically become a reality if testing services (or comparable capabilities) are more fully utilized. The concentration and organization of human and machine resources which a testing service can accomplish, constitutes the basis of most professional examination programs planned for national use.